Class 20: Conditioning & Moment Generating Functions

Tuesday, November 19, 2019 4:08 PM

- Hw 7 due Tues, Nov 26, 1pm

Recap:

- Conditional Expectation as a random variable:

  Calculate $\mathbb{E}[X \mid Y=y] = g(y)$ for all $y \in \mathbb{R}$,

  then plug in the RV Y to get $\mathbb{E}[X \mid Y] = g(Y)$.

- Tower property / law of total expectation

$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X \mid Y]\big]$$

- Law of total variance:

  Conditional variance $\operatorname{Var}(X \mid Y) = \mathbb{E}[X^2 \mid Y] - \big(\mathbb{E}[X \mid Y]\big)^2$

  satisfies:

$$\operatorname{Var}(X) = \mathbb{E}\big[\operatorname{Var}(X \mid Y)\big] + \operatorname{Var}\big(\mathbb{E}[X \mid Y]\big).$$

Ex: (filtering)

$X \sim \operatorname{Exp}(1)$ and $(Y \mid X=x) \sim N\left(x, \frac{x}{2}\right)$

Exponential rate 1

normal r.v. mean x and var. $\frac{x}{2}$

- $\mathbb{E}[Y \mid X=x] = \mathbb{E}\left[N\left(x, \frac{x}{2}\right)\right] = x$ for all $x \in \mathbb{R}$

  so $\mathbb{E}[Y \mid X] = X$

  Tower property: $\mathbb{E}[Y] = \mathbb{E}\big[\mathbb{E}[Y \mid X]\big] = \mathbb{E}[X] = 1$

  $\mathbb{E}[\operatorname{Exp}(\lambda)] = \frac{1}{\lambda}$

- $\operatorname{Var}(Y \mid X=x) = \operatorname{Var}\left(N\left(x, \frac{x}{2}\right)\right) = \frac{1}{2}x$,

  so $\operatorname{Var}(Y \mid X) = \frac{1}{2}X$.

  Law of total variance:

$$\begin{aligned}
\operatorname{Var}(Y) &= \mathbb{E}\big[\operatorname{Var}(Y \mid X)\big] + \operatorname{Var}\big(\mathbb{E}[Y \mid X]\big) \\
&= \mathbb{E}\left[\tfrac{1}{2}X\right] + \operatorname{Var}(X) \\
&= \tfrac{1}{2} + 1 = \tfrac{3}{2}.
\end{aligned}$$

$\mathbb{E}[\operatorname{Exp}(\lambda)] = \frac{1}{\lambda}$

$\operatorname{Var}(\operatorname{Exp}(\lambda)) = \frac{1}{\lambda^2}$

Sum of a random number of random variables

Setup:

- Let $X_1, X_2, X_3, \dots$ be iid random variables.

  independent & identically distributed

  Suppose $\mathbb{E}[X_i] = m \in \mathbb{R}$.

  and $Var(X_i) = v > 0$.

- Let N be a r.v. taking values in $\{0, 1, 2, \dots\}$.
- Assume N and X_i's are independent.
- Goal: Find the mean / variance of

$$S = \sum_{i=1}^{N} X_i = X_1 + X_2 + \dots + X_N.$$

Convention: $\boxed{\sum_{i=1}^{0} = 0}$

- If N is not random this is fairly straightforward:

$$\begin{cases} \mathbb{E}[S] = \mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathbb{E}[X_i] & \text{by linearity of expectation} \\ \quad = \sum_{i=1}^{N} m = Nm. & \\ Var(S) = Var\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} Var(X_i) & \text{by independence of } X_i\text{'s} \\ \quad = Nv & \end{cases}$$

- In general, when N is random, condition on it.

  For each $n = 0, 1, 2, \dots$

$$\begin{aligned} \mathbb{E}[S \mid N=n] &= \mathbb{E}\left[\sum_{i=1}^{N} X_i \mid N=n\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{n} X_i \mid N=n\right] && \leftarrow \text{we know } N=n \text{ inside this expectation} \\ &= \sum_{i=1}^{n} \mathbb{E}[X_i \mid N=n] && \leftarrow \text{since conditional expectation given a fixed event is linear} \\ &= \sum_{i=1}^{n} \mathbb{E}[X_i] && \leftarrow \text{independence of } N \text{ and } X_i \\ &= nm. \end{aligned}$$

$$= nm .$$

This shows $\mathbb{E}[S|N] = Nm$.

Tower property:

$$\mathbb{E}[S] = \mathbb{E}\left[\mathbb{E}(S|N)\right] = \mathbb{E}[Nm]$$

$$= m\,\mathbb{E}[N] .$$ since m is constant

a.k.a. Wald's identity.

Variance:

For $n = 0, 1, 2, \ldots$

$$\mathrm{Var}(S|N=n) = \mathrm{Var}\left(\sum_{i=1}^{N} X_i \,\middle|\, N=n\right)$$

$$= \mathrm{Var}\left(\sum_{i=1}^{n} X_i \,\middle|\, N=n\right)$$

$$= \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)$$ since X_i's and N are independent

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

$$= nv.$$

Thus $\mathrm{Var}(S|N) = Nv$.

Law of total variance:

$$\mathrm{Var}(S) = \mathbb{E}\left[\mathrm{Var}(S|N)\right] + \mathrm{Var}\left(\mathbb{E}[S|N]\right)$$

$$= \mathbb{E}[Nv] + \mathrm{Var}(Nm)$$

$$= \underbrace{v\,\mathbb{E}[N]}_{\text{"variance per customer"}} + m^2 \underbrace{\mathrm{Var}(N)}_{\text{variance due to randomness of } N}.$$

2nd Wald's identity

$$\boxed{\begin{aligned} &\text{Summarize: } \quad m = \mathbb{E}[X_i], \quad v = \mathrm{Var}(X_i) \\ &\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = m\,\mathbb{E}[N] \\ &\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) = v\,\mathbb{E}[N] + m^2\,\mathrm{Var}(N) \end{aligned}}$$

$$\boxed{\begin{aligned} &\underline{\text{Summarize}}: \quad m = \mathbb{E}[X_i], \quad v = Var(X_i) \\ &\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = m\,\mathbb{E}[N] \\ &Var\left(\sum_{i=1}^{N} X_i\right) = v\,\mathbb{E}[N] + m^2 Var(N) \end{aligned}}$$

Ex: We run a taco truck. We get N customers per day, model $N \sim Poisson(80)$.

Each customer spends a random amount of money, uniformly distributed in $[0, 12]$.

So $X_i \sim Unif[0, 12]$.

Note: $\mathbb{E}[N] = 80$, $Var(N) = 80$.

$\mathbb{E}[X_i] = 6$, $Var(X_i) = \frac{12^2}{12} = 12$.

$$\boxed{Var(Unif[a,b]) = \frac{b^2 - a^2}{12}}$$

So $m = 6$, $v = 12$.

- Expected revenue per day $= \mathbb{E}\left[\sum_{i=1}^{N} X_i\right]$

$$= m\,\mathbb{E}[N] = 6 \cdot 80 = 480.$$

- Variance of daily revenue $= Var\left(\sum_{i=1}^{N} X_i\right)$

$$= v\,\mathbb{E}[N] + m^2 Var(N)$$
$$= 12 \cdot 80 + 6^2 \cdot 80$$
$$= 3840.$$

Moment Generating Functions:

Many ways to describe the distribution of a random variable:

CDF	PDF (continuous)	PMF (discrete)
$\mathbb{P}(X \le x)$	$\frac{d}{dx}\mathbb{P}(X \le x)$	$\mathbb{P}(X = x)$

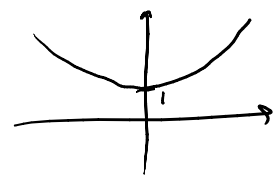
Moment generating function is another way (valid for disc. or cont. RVs).

__Definition:__

- The <u>moment</u> <u>generating</u> <u>function</u> (MGF) of a RV X is the function

$$M_X(t) = E[e^{tX}] \quad \text{for } t \in \mathbb{R}.$$

- Or write just $M(t)$ if we know which X we're working with.



- Note $M_X(0) = \mathbb{E}[e^0] = 1$.

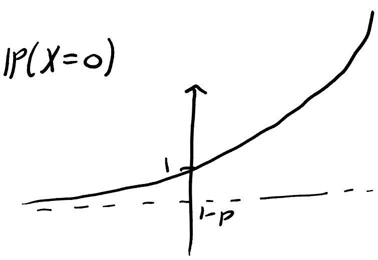
Always convex...

__Ex 1:__ Let $X \sim Ber(p)$. So $X = \begin{cases} 1 & \text{with prob } p \\ 0 & \text{with prob } 1-p \end{cases}$.

Then

$$M_X(t) = \mathbb{E}[e^{tX}] = e^{t\cdot 1}\,\mathbb{P}(X=1) + e^{t\cdot 0}\,\mathbb{P}(X=0)$$

$$= pe^t + 1-p.$$



__Ex 2:__ Let $X \sim Poisson(\lambda)$, $\lambda > 0$.

So $\mathbb{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$.

Then

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{tk}\,\mathbb{P}(X=k)$$

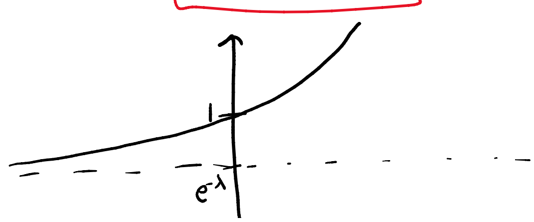
$$= \sum_{k=0}^{\infty} e^{tk}\cdot e^{-\lambda}\cdot\frac{\lambda^k}{k!}$$

$$= e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!}$$

$$\boxed{\text{Taylor series: } e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!}}$$

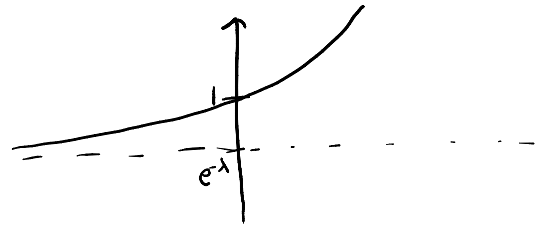
$$= e^{-\lambda}\, e^{\lambda e^t}$$

$$= e^{\lambda(e^t - 1)}$$



$$= e \qquad e$$

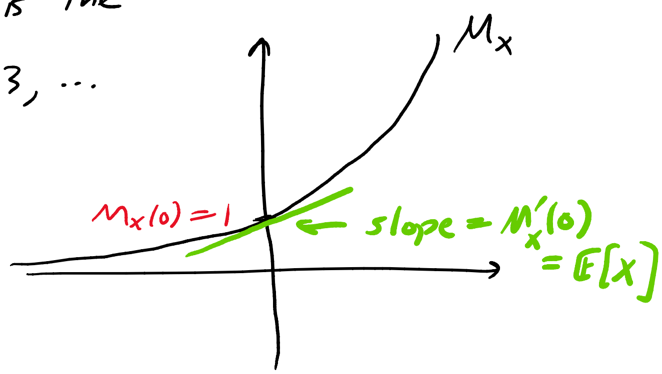
$$= e^{\lambda(e^t - 1)}$$



Def: The <u>n^{th} moment</u> of a RV X is the number $\mathbb{E}[X^n]$, for $n = 1, 2, 3, \ldots$

<u>Ex:</u> $\text{Mean} = \mathbb{E}[X] = 1^{st}$ moment

$\mathbb{E}[X^2] = 2^{nd}$ moment



<u>Fundamental Fact about MGFs:</u>

$$\frac{d^n M_X}{dt^n}(0) = \mathbb{E}[X^n]$$

$\longrightarrow$ n^{th} derivative of $M_X(t)$, evaluated at $t=0$

<u>Proof:</u> Suppose X takes finitely many values.

Then $M_X(t) = \mathbb{E}[e^{tX}] = \sum_k e^{tk} \, \mathbb{P}(X=k)$.

So
$$\frac{dM_X}{dt}(t) = \frac{d}{dt}\left[\sum_k e^{tk} \, \mathbb{P}(X=k)\right]$$
$$= \sum_k \frac{d}{dt}(e^{tk}) \, \mathbb{P}(X=k)$$
$$= \sum_k k e^{tk} \, \mathbb{P}(X=k)$$
$$= \mathbb{E}[X e^{tX}].$$

Thus $\frac{dM_X}{dt}(0) = \mathbb{E}[X e^0] = \mathbb{E}[X]$.

Similarly
$$\frac{d^n M_X}{dt^n}(t) = \sum_k \frac{d^n}{dt^n}(e^{tk}) \, \mathbb{P}(X=k)$$
$$= \sum_k k^n e^{tk} \, \mathbb{P}(X=k)$$
$$= \mathbb{E}[X^n e^{tX}]$$

$$= \sum_k k^n e^{tk}\,\mathbb{P}(X=k)$$

$$= \mathbb{E}[X^n e^{tX}].$$

So $\frac{d^n M_X}{dt^n}(0) = \mathbb{E}[X^n]$.

For a general RV X (not necessarily discrete), same idea, just need to justify:

$$\frac{d^n}{dt^n}\mathbb{E}[e^{tX}] = \mathbb{E}\left[\frac{d^n}{dt^n} e^{tX}\right].$$

<u>Ex:</u> Let $X \sim \mathrm{Exp}(\lambda)$, $\lambda > 0$.

So X has PDF $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$

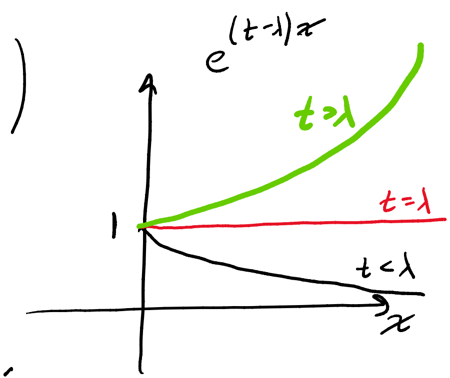
Then

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

$$= \int_0^\infty e^{tx} \cdot \lambda e^{-\lambda x}\,dx = \lambda \int_0^\infty e^{(t-\lambda)x}\,dx$$

$$= \frac{\lambda}{t-\lambda} e^{(t-\lambda)x}\Big|_{x=0}^{\infty} = \frac{\lambda}{t-\lambda}(0-1)$$

$$= \frac{\lambda}{\lambda - t} \qquad \text{if } t < \lambda.$$



But $= \infty$ (integral diverges) if $t \geq \lambda$.

So $$M_X(t) = \begin{cases} \frac{\lambda}{\lambda - t} & \text{if } t < \lambda \\ \infty & \text{if } t \geq \lambda. \end{cases}$$

(Don't worry about the ∞, this is totally ok.)

Let's find some moments: For $t < \lambda$:

$$M_X'(t) = \frac{d}{dt}\left(\frac{\lambda}{\lambda - t}\right) = \frac{\lambda}{(\lambda-t)^2}$$

$$M_X''(t) = \frac{d}{dt}\left(\frac{\lambda}{(\lambda-t)^2}\right) = 2\,\frac{\lambda}{(\lambda-t)^3}$$

$$M_X'''(t) = 2\frac{d}{dt}\left(\frac{\lambda}{(\lambda-t)^3}\right) = 2\left[\,3\,\frac{\lambda}{(\lambda-t)^4}\right] = \frac{6\lambda}{(\lambda-t)^4}$$

$$\vdots$$

$$\frac{d^n M_X}{dt^n}(t) = \frac{n!\,\lambda}{(\lambda-t)^{n+1}}.$$

Thus:

$$\mathbb{E}[X] = \frac{dM_X}{dt}(0) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}. \qquad = \int_0^\infty x \cdot \lambda e^{-\lambda x}\,dx$$

$$\mathbb{E}[X^2] = \frac{d^2M_X}{dt^2}(0) = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2} \qquad = \int_0^\infty x^2 \cdot \lambda e^{-\lambda x}\,dx$$

$$\mathbb{E}[X^n] = \frac{d^nM_X}{dt^n}(0) = \frac{n!\,\lambda}{\lambda^{n+1}} = n!\,\lambda^{-n} \qquad = \int_0^\infty x^n \cdot \lambda e^{-\lambda x}\,dx$$